

Minireview

The diversity of globin-coupled sensors

Tracey Allen K. Freitas^{a,b}, Shaobin Hou^a, Maqsdul Alam^{a,b,*}^aDepartment of Microbiology, Snyder Hall 207, 2538 The Mall; University of Hawaii, Honolulu, HI 96822, USA^bMaui High Performance Computing Center, 550 Lipoa Parkway, Kihei, Maui, HI 96753, USA

Received 2 July 2003; revised 5 August 2003; accepted 10 August 2003

First published online 5 September 2003

Edited by Horst Feldmann

Abstract The recently discovered globin-coupled sensors (GCSs) are heme-containing two-domain transducers distinct from the PAS domain superfamily. We have identified an additional 22 GCSs with varying multi-domain C-terminal transmitters through a search of the complete and incomplete microbial genome datasets. The GCS superfamily is composed of two major subfamilies: the aerotactic and gene regulators. We postulate the existence of protoglobins in Archaea as the predecessor to the chimeric GCS.

© 2003 Published by Elsevier B.V. on behalf of the Federation of European Biochemical Societies.

Key words: Globin; Archaeon; Heme-based sensor; Globin-coupled sensor

1. Introduction

Homo- and heteromeric heme-based sensors are mediators of cellular responses to metabolic and environmental stimuli such as NO, CO and O₂ [1]. Changes in intracellular gas concentrations are sensed by a heme moiety and result in either aerotaxis or gene regulation. Presently, there are six known types of heme sensors: CoxA, NPAS2, sGC, Dos and AXPDEA1, FixL, and HemAT. The HemATs, by homology, are the only aerotactic heme sensors combining globin and MCP signaling domains, whereas the remaining function in gene regulation, either by binding DNA directly, modulating a small metabolite 2nd messenger (cyclic mono- and dinucleotides), or directly interacting with a transcription factor or regulator.

CoxA is a CO sensor that controls the transcription of CO-utilizing genes. Binding of CO to the heme domain of CoxA homodimers modulates the DNA-binding C-terminal domain [2]. Neuronal PAS domain protein 2 (NPAS2) is expressed in mammalian brain tissue [3] and regulates transcription as a heterodimer with BMAL1 [4–6]. Dissociation of the NPAS2:BMAL1 heterodimer occurs upon CO binding to the NPAS2 monomer, effectively removing its DNA-binding, and hence, transcription capability [3]. The soluble guanylate cyclase (sGC) contains a heme-binding and guanylate cyclase domain. Binding of NO to the sGC heterodimer produces cGMP from GTP [7], whereby gene regulation ensues by the

cGMP 2nd messenger. The direct oxygen sensor (Dos), first described in *Escherichia coli* [8], functions as a tetrameric phosphodiesterase (PDE) by converting cAMP to 5'-AMP while in the ferrous form, and is strongly inhibited by CO and NO ligands [9]. A1 from *Acetobacter xylinum* (AxPDEA1) also functions as a PDE by linearizing cyclic bis(3'→5')diguanilate, an allosteric activator of the bacterial cellulose synthase, to the ineffectual pGpG [10,11]. Both Dos and AxPDEA1 possess similar heme-binding PAS domains fused to the PDE C-terminus, consisting of a GGDEF and EAL domain. Histidine kinase FixL binds heme at an N-terminal PAS domain and controls transcription of oxygen-sensitive genes by its response regulator, FixJ [13,14]. Phosphorylated FixJ acts as the transcriptional activator and permits transcription of the *fix* genes [15,16].

Heme-based aerotaxis transducers, the HemATs, possess a heme-binding globin domain and a signaling domain typical of methyl-accepting chemotaxis proteins (MCP) [17]. HemATs, originally discovered in the archaeon *Halobacterium salinarum* and the Firmicutes *Bacillus subtilis*, are members of the family of globin-coupled sensors (GCSs) [18,19]. Variance in the C-terminal transmitter domain indicates that not all GCSs are involved in aerotaxis. In this report, we further identify the diversity of these GCSs resulting from exhaustive searches of completed and in-progress microbial genomes. We also report their putative functions and categorize them in relation to other non-globin heme-based sensors and propose two possible evolutionary models of the GCS and globin.

2. Materials and methods

2.1. Genome and protein sequences

The following preliminary sequence data was obtained from the Institute for Genomic Research website: *Acidithiobacillus ferrooxidans*, *Bacillus anthracis*, *Bacillus cereus*, *Carboxydothermus hydrogenoformans*, and *Geobacter sulfurreducens*; DOE Joint Genome Institute: *Azotobacter vinelandii*, *Burkholderia fungorum*, *Geobacter metallireducens*, *Magnetococcus*, *Magnetospirillum magnetotacticum*, *Rhodobacter sphaeroides*, *Rhodospirillum rubrum*, and *Novosphingobium aromaticivorans*; National Center for Biotechnology Information: *Escherichia coli* O157 H7, *Halobacterium salinarum*, *Agrobacterium tumefaciens*, *Caulobacter crescentus*, *Bacillus halodurans*, *Bacillus subtilis*, *Vibrio vulnificus*, and *Shigella flexneri*; the *Bordetella pertussis*, *Bordetella parapertussis*, and *Bordetella bronchiseptica* sequence data was produced by the *Bordetella pertussis* Sequencing Group at the Sanger Institute and can

*Corresponding author. Fax: (1)-808-956 5339.
E-mail address: alam@hawaii.edu (M. Alam).

be obtained from <ftp://ftp.sanger.ac.uk/pub/pathogens/bpl>. At present, five genomes are incompletely sequenced and therefore accession numbers are not available for those proteins (see Table 1 for details).

2.2. Multiple alignments and secondary structure

All sequences were aligned in a two-stage process. Multiple alignments in ClustalX v1.8 [20] were followed by manual adjustment in DNASTar's MegAlign. At this stage, globin crystal structures (*E. coli* HMP, PDB ID: 1GVH; *Vitreoscilla stercoraria* Hb, PDB ID: 1VHB; *Ralstonia eutropha* FHB,

PDB ID: 1CQX; *Chlamydomonas eugametos* trHb, PDB ID: 1DLY; *Paramecium caudatum* trHb, PDB ID: 1DLW; HemAT-Bs, PDB ID: 1OR6) and Jnet [21] secondary structure predictions were used as guides to produce the finished alignments in Fig. 1A.

2.3. Protein domain detection and analyses

Protein sequences were analyzed with the Pfam (<http://pfam.wustl.edu/>), SMART (<http://smart.embl-heidelberg.de/>), and SCOP (<http://scop.berkeley.edu/>) datasets and domain descriptions were taken from the InterPro database ([**A**

HemAT-Bs Structure

2 helix 3 helix 4 helix 5 helix 6 helix

70 80 90 100 110 120 130 140 150

100 110 120 130 140 150

160 170 180 190 200 210 220 230

240 250 260 270 280 290 300 310 320

330 340 350 360 370 380 390 400 410

420 430 440 450 460 470 480 490 500

510 520 530 540 550 560 570 580 590

600 610 620 630 640 650 660 670 680

690 700 710 720 730 740 750 760 770

780 790 800 810 820 830 840 850 860

870 880 890 900 910 920 930 940 950

960 970 980 990 1000 1010 1020 1030 1040

1050 1060 1070 1080 1090 1100 1110 1120 1130

1140 1150 1160 1170 1180 1190 1200 1210 1220

1230 1240 1250 1260 1270 1280 1290 1300 1310

1320 1330 1340 1350 1360 1370 1380 1390 1400

1410 1420 1430 1440 1450 1460 1470 1480 1490

1500 1510 1520 1530 1540 1550 1560 1570 1580

1590 1600 1610 1620 1630 1640 1650 1660 1670

1680 1690 1700 1710 1720 1730 1740 1750 1760

1770 1780 1790 1800 1810 1820 1830 1840 1850

1860 1870 1880 1890 1900 1910 1920 1930 1940

1950 1960 1970 1980 1990 2000 2010 2020 2030

2040 2050 2060 2070 2080 2090 2100 2110 2120

2130 2140 2150 2160 2170 2180 2190 2200 2210

2220 2230 2240 2250 2260 2270 2280 2290 2300

2310 2320 2330 2340 2350 2360 2370 2380 2390

2400 2410 2420 2430 2440 2450 2460 2470 2480

2490 2500 2510 2520 2530 2540 2550 2560 2570

2580 2590 2600 2610 2620 2630 2640 2650 2660

2670 2680 2690 2700 2710 2720 2730 2740 2750

2760 2770 2780 2790 2800 2810 2820 2830 2840

2850 2860 2870 2880 2890 2900 2910 2920 2930

2940 2950 2960 2970 2980 2990 3000 3010 3020

3030 3040 3050 3060 3070 3080 3090 3100 3110

3120 3130 3140 3150 3160 3170 3180 3190 3200

3210 3220 3230 3240 3250 3260 3270 3280 3290

3300 3310 3320 3330 3340 3350 3360 3370 3380

3390 3400 3410 3420 3430 3440 3450 3460 3470

3480 3490 3500 3510 3520 3530 3540 3550 3560

3570 3580 3590 3600 3610 3620 3630 3640 3650

3660 3670 3680 3690 3700 3710 3720 3730 3740

3750 3760 3770 3780 3790 3800 3810 3820 3830

3840 3850 3860 3870 3880 3890 3900 3910 3920

3930 3940 3950 3960 3970 3980 3990 4000 4010

4020 4030 4040 4050 4060 4070 4080 4090 4100

4110 4120 4130 4140 4150 4160 4170 4180 4190

4200 4210 4220 4230 4240 4250 4260 4270 4280

4290 4300 4310 4320 4330 4340 4350 4360 4370

4380 4390 4400 4410 4420 4430 4440 4450 4460

4470 4480 4490 4500 4510 4520 4530 4540 4550

4560 4570 4580 4590 4600 4610 4620 4630 4640

4650 4660 4670 4680 4690 4700 4710 4720 4730

4740 4750 4760 4770 4780 4790 4800 4810 4820

4830 4840 4850 4860 4870 4880 4890 4900 4910

4920 4930 4940 4950 4960 4970 4980 4990 5000

5010 5020 5030 5040 5050 5060 5070 5080 5090

5100 5110 5120 5130 5140 5150 5160 5170 5180

5190 5200 5210 5220 5230 5240 5250 5260 5270

5280 5290 5300 5310 5320 5330 5340 5350 5360

5370 5380 5390 5400 5410 5420 5430 5440 5450

5460 5470 5480 5490 5500 5510 5520 5530 5540

5550 5560 5570 5580 5590 5600 5610 5620 5630

5640 5650 5660 5670 5680 5690 5700 5710 5720

5730 5740 5750 5760 5770 5780 5790 5800 5810

5820 5830 5840 5850 5860 5870 5880 5890 5900

5910 5920 5930 5940 5950 5960 5970 5980 5990

6000 6010 6020 6030 6040 6050 6060 6070 6080

6090 6100 6110 6120 6130 6140 6150 6160 6170

6180 6190 6200 6210 6220 6230 6240 6250 6260

6270 6280 6290 6300 6310 6320 6330 6340 6350

6360 6370 6380 6390 6400 6410 6420 6430 6440

6450 6460 6470 6480 6490 6500 6510 6520 6530

6540 6550 6560 6570 6580 6590 6600 6610 6620

6630 6640 6650 6660 6670 6680 6690 6700 6710

6720 6730 6740 6750 6760 6770 6780 6790 6800

6810 6820 6830 6840 6850 6860 6870 6880 6890

6900 6910 6920 6930 6940 6950 6960 6970 6980

6990 7000 7010 7020 7030 7040 7050 7060 7070

7080 7090 7100 7110 7120 7130 7140 7150 7160

7170 7180 7190 7200 7210 7220 7230 7240 7250

7260 7270 7280 7290 7300 7310 7320 7330 7340

7350 7360 7370 7380 7390 7400 7410 7420 7430

7440 7450 7460 7470 7480 7490 7500 7510 7520

7530 7540 7550 7560 7570 7580 7590 7600 7610

7620 7630 7640 7650 7660 7670 7680 7690 7700

7710 7720 7730 7740 7750 7760 7770 7780 7790

7800 7810 7820 7830 7840 7850 7860 7870 7880

7890 7900 7910 7920 7930 7940 7950 7960 7970

7980 7990 8000 8010 8020 8030 8040 8050 8060

8070 8080 8090 8100 8110 8120 8130 8140 8150

8160 8170 8180 8190 8200 8210 8220 8230 8240

8250 8260 8270 8280 8290 8300 8310 8320 8330

8340 8350 8360 8370 8380 8390 8400 8410 8420

8430 8440 8450 8460 8470 8480 8490 8500 8510

8520 8530 8540 8550 8560 8570 8580 8590 8600

8610 8620 8630 8640 8650 8660 8670 8680 8690

8700 8710 8720 8730 8740 8750 8760 8770 8780

8790 8800 8810 8820 8830 8840 8850 8860 8870

8880 8890 8900 8910 8920 8930 8940 8950 8960

8970 8980 8990 9000 9010 9020 9030 9040 9050

9060 9070 9080 9090 9100 9110 9120 9130 9140

9150 9160 9170 9180 9190 9200 9210 9220 9230

9240 9250 9260 9270 9280 9290 9300 9310 9320

9330 9340 9350 9360 9370 9380 9390 9400 9410

9420 9430 9440 9450 9460 9470 9480 9490 9500

9510 9520 9530 9540 9550 9560 9570 9580 9590

9600 9610 9620 9630 9640 9650 9660 9670 9680

9690 9700 9710 9720 9730 9740 9750 9760 9770

9780 9790 9800 9810 9820 9830 9840 9850 9860

9870 9880 9890 9900 9910 9920 9930 9940 9950

9960 9970 9980 9990 10000 10010 10020 10030 10040

10050 10060 10070 10080 10090 10100 10110 10120 10130

10140 10150 10160 10170 10180 10190 10200 10210 10220

10230 10240 10250 10260 10270 10280 10290 10300 10310

10320 10330 10340 10350 10360 10370 10380 10390 10400

10410 10420 10430 10440 10450 10460 10470 10480 10490

10500 10510 10520 10530 10540 10550 10560 10570 10580

10590 10600 10610 10620 10630 10640 10650 10660 10670

10680 10690 10700 10710 10720 10730 10740 10750 10760

10770 10780 10790 10800 10810 10820 10830 10840 10850

10860 10870 10880 10890 10900 10910 10920 10930 10940

10950 10960 10970 10980 10990 11000 11010 11020 11030

11040 11050 11060 11070 11080 11090 11100 11110 11120

11130 11140 11150 11160 11170 11180 11190 11200 11210

11220 11230 11240 11250 11260 11270 11280 11290 11300

11310 11320 11330 11340 11350 11360 11370 11380 11390

11400 11410 11420 11430 11440 11450 11460 11470 11480

11490 11500 11510 11520 11530 11540 11550 11560 11570

11580 11590 11600 11610 11620 11630 11640 11650 11660

11670 11680 11690 11700 11710 11720 11730 11740 11750

11760 11770 11780 11790 11800 11810 11820 11830 11840

11850 11860 11870 11880 11890 11900 11910 11920 11930

11940 11950 11960 11970 11980 11990 12000 12010 12020

12030 12040 12050 12060 12070 12080 12090 12100 12110

12120 12130 12140 12150 12160 12170 12180 12190 12200

12210 12220 12230 12240 12250 12260 12270 12280 12290

12300 12310 12320 12330 12340 12350 12360 12370 12380

12390 12400 12410 12420 12430 12440 12450 12460 12470

12480 12490 12500 12510 12520 12530 12540 12550 12560

12570 12580 12590 12600 12610 12620 12630 12640 12650

12660 12670 12680 12690 12700 12710 12720 12730 12740

12750 12760 12770 12780 12790 12800 12810 12820 12830

12840 12850 12860 12870 12880 12890 12900 12910 12920

12930 12940 12950 12960 12970 12980 12990 13000 13010

13020 13030 13040 13050 13060 13070 13080 13090 13100

13110 13120 13130 13140 13150 13160 13170 13180 13190

13200 13210 13220 13230 13240 13250 13260 13270 13280

13290 13300 13310 13320 13330 13340 13350 13360 13370

13380 13390 13400 13410 13420 13430 13440 13450 13460

13470 13480 13490 13500 13510 13520 13530 13540 13550

13560 13570 13580 13590 13600 13610 13620 13630 13640

13650 13660 13670 13680 13690 13700 13710 13720 13730

13740 13750 13760 13770 13780 13790 13800 13810 13820

13830 13840 13850 13860 13870 13880 13890 13900 13910

13920 13930 13940 13950 13960 13970 13980 13990 14000

14010 14020 14030 14040 14050 14060 14070 14080 14090

14100 14110 14120 14130 14140 14150 14160 14170 14180

14190 14200 14210 14220 14230 14240 14250 14260 14270

14280 14290 14300 14310 14320 14330 14340 14350 14360

14370 14380 14390 14400 14410 14420 14430 14440 14450

14460 14470 14480 14490 14500 14510 14520 14530 14540

14550 14560 14570 14580 14590 14600 14610 14620 14630

14640 14650 14660 14670 14680 14690 14700 14710 14720

14730 14740 14750 14760 14770 14780 14790 14800 14810

14820 14830 14840 14850 14860 14870 14880 14890 14900

14910 14920 14930 14940 14950 14960 14970 14980 14990

15000 15010 15020 15030 15040 15050 15060 15070 15080

15090 15100 15110 15120 15130 15140 15150 15160 15170

15180 15190 15200 15210 15220 15230 15240 15250 15260

15270 15280 15290 15300 15310 15320 15330 15340 15350

15360 15370 15380 15390 15400 15410 15420 15430 15440

15450 15460 15470 15480 15490 15500 15510 15520 15530

15540 15550 15560 15570 15580 15590 15600 15610 15620

15630 15640 15650 15660 15670 15680 15690 15700 15710

15720 15730 15740 15750 15760 15770 15780 15790 15800

15810 15820 15830 15840 15850 15860 15870 15880 15890

15900 15910 15920 15930 15940 15950 15960 15970 15980

15990 16000 16010 16020 16030 16040 16050 16060 16070

16080 16090 16100 16110 16120 16130 16140 16150 16160

16170 16180 16190 16200 16210 16220 16230 16240 16250

16260 16270 16280 16290 16300 16310 16320 16330 16340

16350 16360 16370 16380 16390 16400 16410 16420 16430

16440 16450 16460 16470 16480 16490 16500 16510 16520

16530 16540 16550 16560 16570 16580 16590 16600 16610

16620 16630 16640 16650 16660 16670 16680 16690 16700

16710 16720 16730 16740 16750 16760 16770 16780 16790

16800 16810 16820 16830 16840 16850 16860 16870 16880

16890 16900 16910 16920 16930 16940 16950 16960 16970

16980 16990 17000 17010 17020 17030 17040 17050 17060

17070 17080 17090 17100 17110 17120 17130 17140 17150

17160 17170 17180 17190 17200 17210 17220 17230 17240

17250 17260 17270 17280 17290 17300 17310 17320 17330

17340 17350 17360 17370 17380 17390 17400 17410 17420

17430 17440 17450 17460 17470 17480 17490 17500 17510

17520 17530 17540 17550 17560 17570 17580 17590 17600

17610 17620 17630 17640 17650 17660 17670 17680 17690

17700 17710 17720 17730 17740 17750 17760 17770 17780

17790 17800 17810 17820 17830 17840 17850 17860 17870

17880 17890 17900 17910 17920 17930 17940 17950 17960

17970 17980 17990 18000 18010 18020 18030 18040 18050

18060 18070 18080 18090 18100 18110 18120 18130 18140

18150 18160 18170 18180 18190 18200 18210 18220 18230

18240 18250 18260 18270 18280 18290 18300 18310 18320

18330 18340 18350 18360 18370 18380 18390 18400 18410

18420 18430 18440 18450 18460 18470 18480 18490 18500

18510 18520 18530 18540 18550 18560 18570 18580 18590

18600 18610 18620 18630 18640 18650 18660 18670 18680

18690 18700 18710 18720 18730 18740 18750 18760 18770

18780 18790 18800 18810 18820 18830 18840 18850 18860

18870 18880 18890 18900 18910 18920 18930 18940 18950

18960 18970 18980 18990 19000 19010 19020 19030 19040

19050 19060 19070 19080 19090 19100 19110 19120 19130

19140 19150 19160 19170 19180 19190 19200 19210 19220

19230 19240 19250](http://</p>
</div>
<div data-bbox=)

www.ebi.ac.uk/interpro). Various BLAST and PSI-BLAST searches were performed against the non-redundant database and the microbial database at the National Center for Biotechnology Information (<http://www.ncbi.nih.gov/BLAST/>). Transmembrane regions were identified by the algorithms TMHMM2 and DAS (<http://www.cbs.dtu.dk/services/TMHMM-2.0/> and <http://www.sbc.su.se/~miklos/DAS/>).

2.4. Phylogenetic analyses

The distance tree was created using the neighbor-joining (ClustalX) method. Bootstraps (10000 replicates) were calculated directly in ClustalX. Trees were generated in TreeView and NJPlot (distributed with the ClustalX package) and further refined in Adobe Illustrator 10.

3. Results and discussion

An exhaustive heuristic search of the non-redundant protein database and (un)finished microbial genome database at NCBI yielded 27 GCSs. Criteria for identifying a putative GCS included a primary match with the globin domain followed by an accompanying transmitter domain(s). In addition, the length of the globin domain was taken into account as well as the presence of a proximal histidine. In almost all cases, a hydrophobic aromatic residue pair at the end of the B helix (usually Phe-Tyr) was also present. Secondary structure predicting algorithms and the 3D-PSSM fold-recognition server were used to support their inclusion into the family. Using (PSI)BLAST as the primary search algorithm, once a GCS was identified, it was added to the seed alignment. Since the GCS globin domains are highly divergent, each GCS sequence

added to the growing alignment used as a (PSI)BLAST probe for additional candidates.

Neither the SMART database nor the manually curated Pfam-A dataset recognizes the GCS globin domain yet, though the automatically generated Pfam-B family 7730 has an incomplete and partially incorrect (on the basis of the above criteria) GCS globin domain dataset. Fig. 1A represents the alignment of the globin domain of all 27 GCSs. The resulting Neighbor-joining phylogenetic tree was created based on this alignment and is presented in Fig. 1B.

3.1. Biological heme-sensor classification

Using the identified functions of CooA, NPAS2, sGC, Dos, AxPDEA1, FixL, and HemATs, all currently identified biological heme-based sensors can be classified as either aerotactic or gene regulating. Gene regulation is observed to occur via one of three different pathways: via protein–DNA interaction [2–6], via modulation of small-metabolite 2nd messengers [7–12], or by protein–protein interaction as in a transcription factor or regulator [13–16]. The resulting organization schema is illustrated in Fig. 2. GCSs are found in organisms with various physiological and metabolic systems: Gram-positive and Gram-negative, aerobic and anaerobic, oxic and anoxic phototrophs, and even a nitrogen fixer (*A. vinelandii*).

3.1.1. Aerotactic. HemATs are the only known heme-based aerotaxis sensors [17,18] and approximately half of the predicted GCSs are HemATs. Each possess an N-terminal globin domain and a C-terminal MCP-like domain. The original HemAT signaling domain was classified as an ~MCP [17]; however, additional HemATs exhibit a ~HAMP:MCP module. Such a combination is typical of transmitter regions of methyl-accepting chemotaxis proteins such as the *E. coli* serine receptor, Tsr, and hence these proteins may mediate aerotaxis as well. All HemATs are soluble proteins.

The aerotactic subfamily is predominantly Gram-negative α -Proteobacteria (nine proteins), but also includes the Firmicutes (five proteins) and one Archaea. In particular, the magnetotactic proteobacterium *M. magnetotacticum* possesses two aerotactic transducers, whereas *Magnetococcus* MC-1 cells possess only one. Magnetotaxis has been shown to work in conjunction with aerotaxis [22]. Though only a single Archaeal transducer has been found, this is not surprising since at least half of the sequenced Archaeal genomes do not contain recognizable taxis genes. Moreover, the representative sample size of the Archaeal genomes (one GCS out of 18 genomes ~6%) is miniscule compared to that of the bacterial genomes (26 GCSs out of 228 genomes ~11%).

3.1.2. Modulation of a 2nd messenger. Proteins possessing the GGDEF domain have been implicated in c-diGMP modulation [23] and eight such proteins were identified in this group, incorporating either the GGDEF domain or a GGDEF:EAL domain pair. Closer inspection of these proteins reveals another highly conserved domain centered between the N-terminal globin sensor and the C-terminal GGDEF domain. This new domain has been designated as ERERQR, after a conserved patch of residues ($\geq 85\%$ of five acidic, seven basic, 32 polar and 25 hydrophobic sites in a primarily α -helical and coiled structure, data not shown). AfGReg2M has the exact C-terminal domain organization as EcDos and AxPDEA1 (~GGDEF:EAL), PDEs that inactivate the 2nd messengers cAMP and c-diGMP, respectively. The GCS from *B. fungorum* (BfGReg) possesses a C-terminal

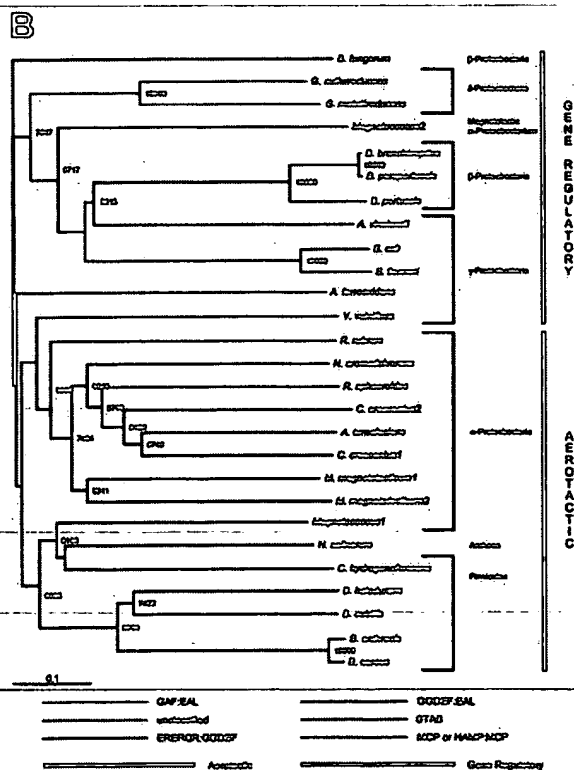


Fig. 1 (Continued).

Table 1
Source information and classification of GCSs

No.	Organism	Name	NCBI accession no.	Classification	SMART	Pfam	Taxonomy	Protein length	Other name
1	<i>Bacillus anthracis</i>	HemAT-Ba	NP_653892	Aerotactic	MA	MCP	Firmicutes	434	BA_0532
2	<i>Bacillus halodurans</i>	HemAT-Bh	NP_241371	Aerotactic	MA	MCP	Firmicutes	441	BH505
3	<i>Bacillus subtilis</i>	HemAT-Bs	NP_388919	Aerotactic	MA	MCP	Firmicutes	433	YhIV
4	<i>Bacillus cereus</i>	HemAT-Bc	NP_835085	Aerotactic	MA	MCP	Firmicutes	434	-
5	<i>Carboxydothermus hydrogenoformans</i>	HemAT-Ch	TIGR_129958	Aerotactic	MA	MCP	Firmicutes	251	-
6	<i>Halobacterium</i> sp. NRC-1	HemAT-Hs	NP_280321	Aerotactic	MA	MCP	Archaea	490	HtrX, HtrB, Htr10
7	<i>Magnetospirillum magnetotacticum</i>	HemAT-MmA	ZP_00054774	Aerotactic	MA	MCP	α-Proteobacteria	444	Magn7582
8	<i>Magnetospirillum magnetotacticum</i>	HemAT-MmB	ZP_00054075	Aerotactic	MA	MCP	α-Proteobacteria	732	Magn6867
9	<i>Rhodobacter sphaeroides</i>	HemAT-Rs	ZP_0006252	Aerotactic	MA	MCP	α-Proteobacteria	371	Rsp2166
10	<i>Rhodospirillum rubrum</i>	HemAT-Rr	ZP_00014161	Aerotactic	MA	MCP	α-Proteobacteria	442	Rrub1164
11	<i>Agrobacterium tumefaciens</i>	HemAT-At	NP_354049	Aerotactic	HAMP-MA	HAMP-MCP	α-Proteobacteria	500	AGR_C_1888
12	<i>Caulobacter crescentus</i>	McpB	NP_419247	Aerotactic	HAMP-MA	HAMP-MCP	α-Proteobacteria	538	McpB
13	<i>Caulobacter crescentus</i>	McpM	NP_421120	Aerotactic	HAMP-MA	HAMP-MCP	α-Proteobacteria	556	McpM
14	<i>Novosphingobium aromaticivorans</i>	HemAT-Na	ZP_00095064	Aerotactic	HAMP-MA	HAMP-MCP	α-Proteobacteria	482	Saro2089
15	<i>Magnetococcus</i> sp. MC-1	HemAT-Mg	ZP_00043038	Aerotactic	HAMP-MA	HAMP-MCP	α-Proteobacteria	519	Mmc10749
16	<i>Magnetococcus</i> sp. MC-1	MgGReg	ZP_00042662	Gene regulator (2nd messenger)	ERERQR:DUF1	ERERQR:GGDEF	α-Proteobacteria	467	Mmc10355
17	<i>Bordetella bronchiseptica</i>	BbGReg	n/a	Gene regulator (2nd messenger)	ERERQR:DUF2	ERERQR:GGDEF	β-Proteobacteria	531	-
18	<i>Bordetella parapertussis</i>	BpaGReg	n/a	Gene regulator (2nd messenger)	ERERQR:DUF3	ERERQR:GGDEF	β-Proteobacteria	514	-
19	<i>Bordetella pertussis</i>	BpeGReg	n/a	Gene regulator (2nd messenger)	ERERQR:DUF4	ERERQR:GGDEF	β-Proteobacteria	531	-
20	<i>Escherichia coli</i>	EcGReg	NP_287665	Gene regulator (2nd messenger)	ERERQR:DUF5	ERERQR:GGDEF	γ-Proteobacteria	460	YddV
21	<i>Azotobacter vinelandii</i>	AvGReg	ZP_00090857	Gene regulator (2nd messenger)	ERERQR:DUF6	ERERQR:GGDEF	γ-Proteobacteria	472	Avin2552
22	<i>Shigella flexneri</i> 2a str.301	SfGReg	NP_707605	Gene regulator (2nd messenger)	ERERQR:DUF7	ERERQR:GGDEF	γ-Proteobacteria	381	YddV
23	<i>Acidithiobacillus ferrooxidans</i>	AfGReg	n/a	Gene regulator (2nd messenger)	DUF1-DUF2	GGDER:EAL	γ-Proteobacteria	880	-
24	<i>Burkholderia fungorum</i>	BfGReg	ZP_00030046	Gene regulator (2nd messenger or Tmscon Reg)	GAF:DUF2	GAF-EAL	β-Proteobacteria	724	Beep2859
25	<i>Vibrio vulnificus</i> CMCP6	VvGReg	NP_762059	Gene regulator (2nd messenger)	STAS	STAS	γ-Proteobacteria	306	VV20073
26	<i>Geobacter sulfurreducens</i>	GsGCS	n/a	Unclassified	-	-	δ-Proteobacteria	300	-
27	<i>Geobacter metallireducens</i>	GmGCS	ZP_00082251.1	Unclassified	-	-	δ-Proteobacteria	300	Gmet3020

Accompanying each GCS is the source organism, suggested naming convention along with any previous names, NCBI accession numbers (available except for those with genome sequencing in-progress), classification according to Fig. 2, domain topology as identified by SMART and Pfam, taxonomy and sequence length. Naming conventions for the GCSs are as follows: HemAT = heme-based aerotactic transducers; GReg = gene regulating.

Biological Heme-based Sensors

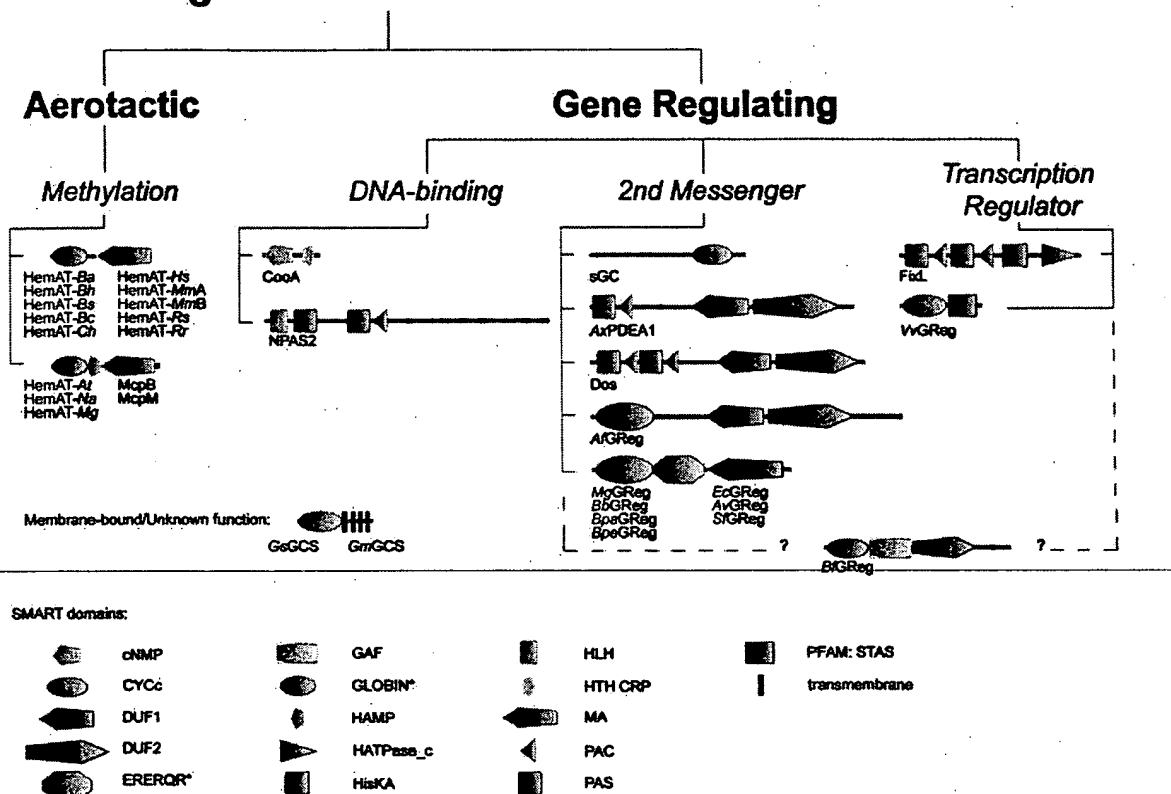


Fig. 2. Functional classification scheme of biological heme-based sensors. Heme-based sensors CooA, NPAS2, sGC, AxPDEA1, Dos, FixL, and HemAT can be grouped according to their primary functions described in the literature. The GCSs are tentatively categorized according to this schema. BfGRReg is believed to be a gene regulator of either the 2nd messenger or transcription regulator class. No function could be assigned to the two membrane-bound *Geobacter* GCSs, GsGCS and GmGCS. Domains with an asterisk (*) indicate new domains not presently a part of the SMART database. See text and Table 1 for details.

GAF:EAL together with an additional PAS domain. Proteins possessing the GAF domain regulate small molecules like cAMP and cGMP and function in transcription [23–25].

3.1.3. Protein–protein interactions. VvGRReg from *V. vulnificus* possesses a C-terminal STAS (sulfate transporter and anti- σ factor antagonist) domain recognized by Pfam as an anti-anti- σ factor. Spore formation in *B. subtilis* is an example of such a regulated process utilizing σ^F (σ factor initiating prespore development), its antagonist SpoIIAB, and the anti- σ factor, SpoIIAA. To our knowledge, VvGRReg is the first example of a globin domain with a transcriptional regulator. GCSs predicted to be involved in DNA binding have yet to be identified.

3.1.4. Unclassified GCS. Two GCSs identified in the strict anaerobic δ -Proteobacteria may be involved in sulfate/sulfur reduction. GsGCS from *G. sulfurreducens* and GmGCS from *G. metallireducens* exhibit a bundle of four transmembrane helices at C-terminal resemble either glutathione *S*-transferase (GST) or ferritin-like proteins. These are generally soluble proteins; however, a distinct microsomal membrane-bound GST family has been identified [26,27]. Both proteins are involved in cellular protection from toxicity of reactive oxygen species [28].

3.2. Phylogenetics of the GCSs

The phylogenetic tree (Fig. 1B) results in two interpreta-

tions: (1) there is a predisposition of bacterial lineages for particular signal-transducing elements, or (2) the globin domains are customized to function in concert with particular signal-transducing elements.

In the case of the GCS, a more evolved and ordered protein is built up from the less ordered components; namely, the ancestor globin, or protoglobin, and the signaling domains. This higher ordered protein imparts a new function(s) to the host organism that allow descendants to thrive in environments that may not have been able to survive before. Rapid response to toxic oxygen or other highly reactive species that otherwise might quickly kill a microbe is a significant pressure to retain such a fusion protein. Within the tenet of the biological evolution, as atmospheric oxygen levels rose and eukaryotic cells evolved, the need for oxygen taxis may have diminished, resulting in the absence of such chimeric systems in the upper eukaryotes. There are three organisms that possess two GCSs: *C. crescentus*, *M. magnetotacticum*, and *Magnetococcus*. All four proteins in *C. crescentus* and *M. magnetotacticum* are HemATs and therefore it seems likely that they arose from gene duplication, i.e. they are paralogs. In contrast, the two GCSs from *Magnetococcus* perform different functions. One is a HemAT and the other, a predicted gene regulator. This indicates that each globin evolved independently with its particular signaling domain to reflect the observed diversity (Fig. 1B) and predicts the existence of the

protoglobin in more primitive organisms like the Archaea or the deeply branching photosynthetic bacteria.

4. Summary

The diversity of heme-based sensors in prokaryotes is predominantly globin based. The family of GCSs can be grouped into two subfamilies, the aerotactic and the gene regulating. Though approximately half of the GCSs fall into the gene-regulating subfamily, the HemATs are the only known heme-based sensors involved in aerotaxis. The GCSs form a family of proteins (Fig. 2) that, thus far, populate all but the direct DNA-binding sensors. Considering the diversity of the GCSs and that the flavohemoglobins are similar to the GCSs, we propose that this form of globin was particularly suited for forming multi-domain chimeric proteins with novel functions. We postulate that protoglobin was the predecessor to the chimeric GCS and should therefore be found in more ancient organisms, like the Archaea.

Acknowledgements: We thank Sergei Vinogradov and an unknown reviewer for critical comments on the manuscript. This investigation was supported by the National Science Foundation grant no. MCB 0080125 and by the University of Hawaii intramural bioinformatic grant awarded to M.A.

References

- [1] Chan, M.K. (2001) *Curr. Opin. Chem. Biol.* 5, 216–222.
- [2] Lanzilotta, W.N., Schuller, D.J., Thorsteinsson, M.V., Kerby, R.L., Roberts, G.P. and Poulos, T.L. (2000) *Nat. Struct. Biol.* 7, 876–880.
- [3] Dioum, E.M., Rutter, J., Tuckerman, J.R., Gonzalez, G., Gilles-Gonzalez, M.A. and McKnight, S.L. (2002) *Science* 298, 2385–2387.
- [4] Gekakis, N., Staknis, D., Nguyen, H.B., Davis, F.C., Wilsbacher, L.D., King, D.P., Takahashi, J.S. and Weitz, C.J. (1998) *Science* 280, 1564–1569.
- [5] Hogenesch, J.B., Gu, Y.Z., Jain, S. and Bradfield, C.A. (1998) *Proc. Natl. Acad. Sci. USA* 95, 5474–5479.
- [6] Reick, M., Garcia, J.A., Dudley, C. and McKnight, S.L. (2001) *Science* 293, 506–509.
- [7] Zhao, Y., Brandish, P.E., Ballou, D.P. and Marletta, M.A. (1999) *Proc. Natl. Acad. Sci. USA* 96, 14753–14758.
- [8] Delgado-Nixon, V.M., Gonzalez, G. and Gilles-Gonzalez, M.A. (2000) *Biochemistry* 39, 2685–2691.
- [9] Sasakura, Y., Hirata, S., Sugiyama, S., Suzuki, S., Taguchi, S., Watanabe, M., Matsui, T., Sagami, I. and Shimizu, T. (2002) *J. Biol. Chem.* 277, 23821–23827.
- [10] Tal, R., Wong, H.C., Calhoon, R., Gelfand, D., Fear, A.L., Volman, G., Mayer, R., Ross, P., Amikam, D., Weinhouse, H., Cohen, A., Sapir, S., Ohana, P. and Benziman, M. (1998) *J. Bacteriol.* 180, 4416–4425.
- [11] Chang, A.L., Tuckerman, J.R., Gonzalez, G., Mayer, R., Weinhouse, H., Volman, G., Amikam, D., Benziman, M. and Gilles-Gonzalez, M.A. (2001) *Biochemistry* 40, 3420–3426.
- [12] Weinhouse, H., Sapir, S., Amikam, D., Shilo, Y., Volman, G., Ohana, P. and Benziman, M. (1997) *FEBS Lett.* 416, 207–211.
- [13] Gilles-Gonzalez, M.A., Ditta, G.S. and Helinski, D.R. (1991) *Nature* 350, 170–172.
- [14] Gilles-Gonzalez, M.A. and Gonzalez, G. (1993) *J. Biol. Chem.* 268, 16293–16297.
- [15] David, M., Daveran, M.L., Batut, J., Dedieu, A., Domergue, O., Ghai, J., Hertig, C., Boistard, P. and Kahn, D. (1988) *Cell* 54, 671–683.
- [16] Virts, E.L., Stanfield, S.W., Helinski, D.R. and Ditta, G.S. (1988) *Proc. Natl. Acad. Sci. USA* 85, 3062–3065.
- [17] Hou, S., Larsen, R.W., Boudko, D., Riley, C.W., Karatan, E., Zimmer, M., Ordal, G.W. and Alam, M. (2000) *Nature* 403, 540–544.
- [18] Hou, S., Belisle, C., Lam, S., Piatibratov, M., Sivozhlezov, V., Takami, H. and Alam, M. (2001) *Extremophiles* 5, 351–354.
- [19] Hou, S., Freitas, T., Larsen, R.W., Piatibratov, M., Sivozhlezov, V., Yamamoto, A., Meleshkevitch, E.A., Zimmer, M., Ordal, G.W. and Alam, M. (2001) *Proc. Natl. Acad. Sci. USA* 98, 9353–9358.
- [20] Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. (1997) *Nucleic Acids Res.* 24, 4876–4882.
- [21] Cuff, J.A. and Barton, G.J. (1999) *Proteins* 40, 502–511.
- [22] Frankel, R.B., Bazylnski, D.A., Johnson, M.S. and Taylor, B.L. (1997) *Biophys. J.* 73, 994–1000.
- [23] Galperin, M.Y., Nikolskaya, A.N. and Koonin, E.V. (2001) *FEMS Microbiol. Lett.* 203, 11–21.
- [24] Kanacher, T., Schultz, A., Linder, J.U. and Schultz, J.E. (2002) *EMBO J.* 21, 3672–3680.
- [25] Hurley, J.H. (2003) *Sci. STKE* 164, PE1, Review.
- [26] Prabhu, K.S., Reddy, P.V., Gumprecht, E., Hildenbrandt, G.R., Scholz, R.W., Sordillo, L.M. and Reddy, C. (2001) *Biochem. J.* 360, 345–354.
- [27] Morgenstern, R., Guthenberg, C. and Depierre, J.W. (1982) *Eur. J. Biochem.* 128, 243–248.
- [28] Raza, H., Robin, M.A., Fang, J.K. and Avadhani, N.G. (2002) *Biochem. J.* 366, 45–55.

BEST AVAILABLE COPY